



Technological University Dublin
ARROW@TU Dublin

Conference papers

School of Computing

2016-09-09

Model-Based and Model-Free Active Learning for Regression

Jack O'Neill

Technological University Dublin, d13128926@mydit.ie

Sarah Jane Delany

Technological University Dublin, sarahjane.delany@tudublin.ie

Brian MacNamee

University College Dublin, brian.macnamee@ucd.ie

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomcon>



Part of the [Computer Engineering Commons](#)

Recommended Citation

O'Neill, J., Delany, S. J. & MacNamee, B. (2016) Model-Based and Model-Free Active Learning for Regression, *16th Annual UK Workshop on Computational Intelligence*, Lancaster, 7th - 9th September, 2016.

This Conference Paper is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#)



Model-Free and Model-Based Active Learning for Regression

Jack O’Neill, Sarah Jane Delany and Brian MacNamee

Abstract Training machine learning models often requires large labelled datasets, which can be both expensive and time-consuming to obtain. *Active learning* aims to selectively choose which data is labelled in order to minimize the total number of labels required to train an effective model. This paper compares model-free and model-based approaches to active learning for regression, finding that model-free approaches, in addition to being less computationally intensive to implement, are more effective in improving the performance of linear regressions than model-based alternatives.

1 Introduction

Machine Learning algorithms are typically used to induce models capable of predicting one or more features (or *target values*) of previously unseen data. Model induction usually relies on the availability of a corpus of *training data*, within which these target values have been previously labelled. It is well recognized that for some tasks, this corpus is “*very difficult, time-consuming, or expensive to obtain*” [1]. Recent years have seen a growing awareness of the cost involved in labelling data [2, 3]. This has led to an increased interest in active learning [1]; a sub-field of machine learning which seeks to generate models that perform better with less data by carefully selecting the data from which the model learns.

In the absence of active learning, an *oracle* — usually a human domain-expert — is presented with a large set of unlabelled data, \mathcal{U} . The oracle provides labels for

Jack O’Neill
Dublin Institute of Technology e-mail: d13128926@mydit.ie

Sarah Jane Delany
Dublin Institute of Technology e-mail: sarahjane.delany@dit.ie

Brian Mac Namee
University College Dublin e-mail: brian.macnamee@ucd.ie

each unlabelled instance, adding it to the set of labelled data, \mathcal{L} , until all data has been labelled. Active Learning, on the other hand, works towards creating this fully-labelled dataset incrementally. An initial, or *seed* batch of data is selected from \mathcal{U} and added to the labelled dataset, \mathcal{L} . A selection strategy, or *query strategy*, is then employed to choose a small subset — or *batch* — of data from \mathcal{U} , which is presented to the oracle for labelling. This batch of data is then labelled and added to \mathcal{L} . The process is repeated until some pre-defined stopping criterion has been reached. If the goal of the exercise is to create a fully labelled dataset, the oracle may only ever be presented with a subset of \mathcal{U} ; a predictive model may then be induced using the resulting labelled data and used to predict the labels of the remaining instances in \mathcal{U} ; saving time and effort on the labelling process.

A key component of the active learning cycle is the query selection strategy. This selection strategy is a technique employed to select unlabelled data to present to the oracle for labelling. Selection strategies ascribe some measure of the predicted *utility* of each unlabelled instance to the effectiveness of the model to be trained. This prediction may be made using information derived from the labelled data already provided, \mathcal{L} , or the remainder of the as-yet unlabelled data, \mathcal{U} . One approach to predicting the utility of an unlabelled instance, which we refer to as *model-based*, is to generate a model using the currently labelled data, \mathcal{L} , and choose unlabelled instances based on information derived from this model. Five of the six query strategy frameworks, or *families*, identified by Settles in his literature survey of Active Learning [1] (uncertainty sampling, Query-by-Committee (QBC), Expected Model Change (EMC), Expected Error Reduction, and Variance Reduction) are model-based approaches. The sixth, Density-weighted, as its name suggests, is generally employed to weight the output of one of the other five frameworks, and is not often used as a standalone approach. Although effective, model-based approaches incur the performance overhead of having to repeatedly generate predictive models, thus increasing the time required to label data. Model-free approaches, on the other hand, predict the informativeness of candidate instances based solely on the non-label features of the datasets \mathcal{U} and \mathcal{L} . Candidate instances may be selected solely on their *density*, or representativeness of the rest of the data, or *diversity* *i.e.* how dissimilar they are from the unlabelled or labelled datasets. This obviates the need for repeated model induction which we would expect to reduce the computational overhead and time required for the selection strategy to generate queries.

This study compares the effectiveness of model-based approaches to that of model-free approaches in the context of dataset labelling using a linear regression model. The current-state-of-the-art in active learning for regression — represented by the QBC and EMC algorithms — is evaluated against three model-free approaches; a purely density-based approach (labelled *Density*), a purely diversity-based approach (labelled *Diversity*) and a mixture of both, realized in the Exploration Guided Active Learning (*EGAL*) algorithm [4].

Section 2 outlines the challenges involved in applying Active Learning to regression models, along with the current approaches taken to overcome these challenges. Section 3 describes the datasets and algorithms used in the evaluation. Section 4

presents the results of the evaluation while Section 5 discusses these findings and suggests future directions of research.

2 Related Work

There is an abundance of literature examining the applicability of active learning to problems of classification [1]. However, the use of active learning for regression has received considerably less attention [5]. The theoretical capability of active learning to significantly improve the rate and accuracy of estimating a function in the presence of noise has been shown by Castro *et al.* [6]. This study shows that the learning rate may be increased when learning functions “*whose complexity is highly concentrated in small regions of space*” [6] *i.e.* functions generally better suited to kernel-based models. This is due to the ability of active learning techniques to quickly isolate *interesting* regions of the version space using techniques such as model uncertainty and local variance. Although the study does not find that active learning could provably significantly outperform passive sampling in learning a general function without localized complexity, the goal of dataset labelling is not necessarily to approximate a general function, applicable to unseen data, but rather to discover that function which best approximates the finite sample of data in the unlabelled pool. Furthermore, earlier work by Sung and Niyogi [7] showed promising practical results in applying active learning to estimating polynomial functions which do not have this property of localized complexity.

Although active learning for classification problems has received more attention than active learning for regression, a number of active learning selection strategies initially developed for use in classification have been shown to perform well when used in regression problems. These approaches, Expected Gradient Length (EGL) and QBC, fall primarily into the category of *model-based* selection strategies. These selection strategies assign an expected utility of each unlabelled instance based on the output of models generated using the currently labelled data. Another approach, *Transductive Experimental Design*, outlined by Yu *et al.* [8] employs statistical techniques from *Optimal Experiment Design* to assess the utility of an instance based on its non-label features. This approach avoids the need to train additional models and reduces the overhead incurred in employing an active learning selection strategy.

QBC, originally introduced by Seung *et al.* [9], is an example of an ensemble-based approach to active learning. QBC trains a committee of models using different *views* of the available data; and selects for labelling the unlabelled instance on which each of the models in this committee most disagree. Burbidge *et al.* [10] have explored the application of the QBC algorithm to linear regression models, finding it to perform favourably against a random baseline.

Expected Gradient Length was proposed for classification problems by Settles and Craven in 2008 [11], and is a variation on QBC. Like QBC, it builds a committee of models using samples of the labelled dataset. However, unlike QBC, unlabelled data is scored on the basis of the disagreement between the aggregated output of

the committee on the one hand, and the predicted outcome of a model built on the entire labelled dataset — the *output model* — on the other. The idea behind EGL is that those instances which maximise the change in the output model are most likely to improve the model’s accuracy. Cai *et al.* [5] have applied EGL (which they also refer to as Expected Model Change) to regression problems, finding it to outperform QBC on a number of datasets.

Diversity, or distance from the labelled data, was first proposed as a sole measurement of the usefulness of an unlabelled instance by Baram *et al.* [12] using their Kernel Farthest-First (KFF) algorithm. The KFF algorithm is based, as its name suggests, on farthest-first traversal sequences in kernel space. The KFF algorithm seeks to label the unlabelled instance which is least similar to (i.e. farthest from) the currently labelled dataset, with the distance between a point and a set defined as the minimum distance between that point and any instance belonging to the set. The KFF algorithm has been shown to outperform a random baseline on a simple XOR classification problem [12], and has been incorporated in the work of Osugi *et al.* [13]; but has received little attention as a stand-alone approach to active learning.

Density, or *closeness* to the labelled data, was proposed as a selection strategy when comparing multiple algorithms by Boiy and Moens [14]. Their density-based selection strategy was implemented as the inverse of the Kernel Farthest-First Diversity algorithm described above — labelling those instances closest to the currently labelled set. Density-based selection strategies are not generally as effective as the other approaches described above; the problems with density-based selection strategies have been explained and visualized by MacNamee *et al.* [15]. Both Density and Diversity have been implemented in accordance with the approach taken by Hu *et al.* [16], using cosine distance as a similarity measure between instances.

Exploration Guided Active Learning (EGAL) was proposed by Hu *et al.* [4] as a classifier-independent approach which offers computational advantages over committee-based alternatives. The EGAL algorithm is a model-free approach to active learning based on a combination of density and diversity measures. Unlabelled data is compared using a similarity measure (cosine similarity was chosen by Hu *et al.* [16], but the approach is independent of the particular measure used), and only those instances which are sufficiently *distant* from the currently labelled dataset — the *candidate set* — are considered for labelling. Within this candidate set, instances are ranked according to their *density* within the dataset as a whole, and those instances with the greater density are preferred. EGAL works on the assumption that the densest instances are most representative of the current data, allowing EGAL to balance a bias for selecting dissimilar instances for labelling with a resilience to labelling outliers which are not representative of the data as a whole.

A random selection strategy is a model-free selection strategy which is often used as a baseline against which to compare proposed active learning approaches [10, 5, 4]. This approach simply selects unlabelled data at random and makes no attempt to maximize the utility of the queried labels; making it a suitable simulation of passive learning.

Many active learning experiments are conducted by allowing the selection strategy to query the label of a single instance for each iteration of the active learning

loop, which is not always a realistic setting [17]. A lower batch size allows a selection strategy to make a larger number of queries, each of which is built upon a larger collection of prior labels. However, this also increases the number of models which the selection strategy has to generate, and by extension, the time taken to complete the active learning loop.

While numerous studies have been undertaken aiming to optimize batch construction [18, 19, 20], the impact of increasing batch size on the learning rate of selection strategies has not been studied in detail.

Although active learning is often used to minimize the number of labelled examples required to induce effective generalizable models, previous work on active learning [21, 22] has shown its applicability to dataset labelling, a related, but subtly different task. In the former, models are induced in order to make predictions about previously unseen data. When applied to dataset labelling the goals of the induced model are to “*minimise the number of labels that are needed from the human expert while maintaining the quality of the labels applied by the system*” [21]. The resulting labelled data may in turn be used to induce a model which *would* be required to generalize, but this is not a concern during the labelling phase.

3 Experimental Framework

The aim of this evaluation is to compare the performance of a broad range of active learning strategies in the context of regression-based tasks. Each of the active learning selection strategies under consideration was tested on seven separate datasets, all of which are available from the UCI Machine Learning Repository¹. The datasets used are summarized in Table 1, which lists the number of features and instances of each dataset following preparatory treatment outlined below.

Datasets were chosen on the basis of their suitability for inducing a regression model, however, some pre-processing was required on the *bike sharing demand* dataset before proceeding. This dataset describes daily demand for an urban bike sharing service. The *instant* and *dteday* features, an ID and date identifier respectively, were dropped as they are both necessarily unique to each instance. The features, *season*, *mnth*, *yr* and *weekday* are more properly categorical features. As linear regressions are unable to process categorical features, these were expressed using one-hot encoding. The features *casual* and *registered* simply break down the target feature *cnt* into number of registered members and number of unregistered members availing of the bike sharing service. These were removed as they allow for simple direct computation of the label. A working R script demonstrating the required processing is available online²

All other datasets were processed as is. Table 1 gives a brief summary of each dataset, including the number of features, number of instances and the *r* squared

¹ <https://archive.ics.uci.edu/ml/datasets.html>

² link removed for anonymization purposes

value of the linear regression created by using all of the available data. This value represents the correlation between the label and the rest of the features for each dataset.

Table 1 Description of Datasets Used

Dataset	Features	Instances	R Squared
Auto MPG	8	392	0.82
House Prices	14	506	0.74
Treasury	16	1049	1.00
Wine Quality Red	12	1599	0.36
Wine Quality White	12	4898	0.28
Concrete Data	9	1030	0.62
Bike Sharing	34	731	0.84

All of the datasets used were fully labelled. This made it possible to simulate the labelling process without the need for a human oracle; as labels were *hidden* from the system until requested. The experiment proceeded as follows. An initial batch of 10 instances was randomly selected and added to the labelled set \mathcal{L} as seed data. Execution continued in iterations. At each iteration, both the labelled set, \mathcal{L} and unlabelled set, \mathcal{U} , were presented to the Active Learning Selection Strategy, which generated a query requesting \mathcal{B} labels. In order to investigate the impact of the batch size, \mathcal{B} , on the active learning process, the experiment was repeated using a range of batch sizes between 1 and 50. The labels for each of these instances were revealed and added to the labelled set. The labelled data was then used to induce a linear regression model, which provided predicted labels for the remaining unlabelled data \mathcal{U} . The process was then repeated until no unlabelled data remained.

Using the error recorded after each iteration, a learning curve was constructed plotting the error of the linear regression induced after each batch iteration as a function of the number of labels in \mathcal{L} . Performance is reported both graphically (showing the Root Mean Square Error (RMSE) of all algorithms after each batch has been processed), and tabularly; condensing the scores across iterations using trapezoidal integration [23] resulting in an *Area Under the Learning Curve* (AULC) score, representing its overall performance — a lower AULC score indicating a stronger performance.

The EGAL algorithm requires a similarity measure used to calculate density and diversity, an α parameter which controls the neighbourhood size for density calculations, and an ω , or *balancing* parameter [4] — which controls the relative contribution of density and diversity in instance selection. Following Hu *et al.*’s recommendation, the standard EGAL algorithm used an α parameter of the mean of the similarities within the dataset less 0.75 the standard deviation, $(\mu - 0.75\sigma)$, and an ω parameter of 0.25. These parameters have been shown to be generally optimal in problems of classification [16]. The *cosine distance* similarity measure was chosen for both algorithms due to its simplicity and popularity.

The Random selection strategy was used as a baseline against which to compare the other algorithms under scrutiny. In order to obtain a reliable estimate of the performance of the random selection strategy, results were averaged over 10 iterations, each created using a unique random seed. Variance between each instantiation of the Random selection strategy was recorded to allow performance graphs to be fitted with error bars, giving a good indication of the range of performances expected in the absence of an active learning selection strategy.

The QBC and EMC algorithms each rely on measuring *disagreement* between various models. Both QBC and EMC were implemented using committee sizes of 5 after prior testing showed this to perform well. In the case of QBC, disagreement is measured as the sum of the absolute differences between each committee member’s predicted output value, and the average predicted output of the committee as a whole. In the case of EMC, disagreement is measured as the absolute difference between the current model on the one hand and the average output of the 5-member committee on the other.

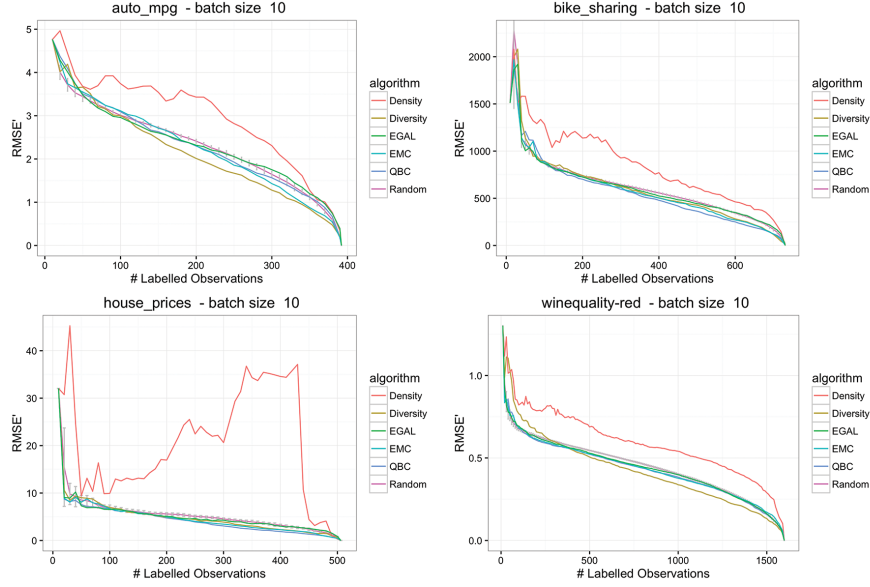
4 Evaluation

Figure 1 shows the learning curve for each algorithm on a selection of datasets using a batch size of 5. It is immediately clear that the Density algorithm performs poorly, having a substantially larger AULC than the random baseline. The difference in performance between the rest of the algorithms is less pronounced. However, the Diversity algorithm occasionally outperforms the alternatives (*auto mpg*, *wine quality red*) and all algorithms apart from Density appear to perform at least as well as the random baseline consistently. The strong performance of Diversity on the one hand, and poor performance of Density on the other, suggest that the EGAL algorithm may have potential for improvement. A balancing parameter of 0.25 was found by Hu *et al.* to be optimal for classification problems [4], these findings suggest that a lower value, favouring diversity over density may be optimal for regression problems.

Table 2 shows the AULC score achieved by each of the selection strategies on all datasets, using a batch size of 5. As the absolute AULC scores achieved are incommensurable between datasets, a Friedman Rank Sum test was performed to detect statistical difference between algorithms. Table 3 shows these performances ranked within datasets.

The Diversity algorithm showed the strongest performance on 4 of the 7 datasets. However, the QBC algorithm achieved the most consistent performance, being among the top 2 highest performing algorithms on each dataset; suggesting that the performance of QBC is more reliable regardless of the dataset used.

A Friedman Test for statistically significant difference between algorithms’ performances yields a p value <0.001 , indicating that at least one of the active learning algorithms under consideration has a statistically significant impact on the data labelling process. A post-hoc Nemenyi test, shown in Table 4, demonstrates that the

Fig. 1 Learning Curves on selected datasets at batch size 5**Table 2** AULC for all algorithms, with the lowest AULC shown in bold

	Density	Diversity	EGAL	EGL	Mean Random	QBC
Auto MPG	1138.67	812.95	895.45	863.45	891.62	838.50
Wine Quality Red	940.57	666.75	696.39	687.06	707.53	686.68
Wine Quality White	3072.79	2461.34	2449.73	2385.70	2479.72	2395.43
Treasury	398.19	149.02	172.75	168.20	187.05	166.91
House Prices	10087.72	2241.96	2380.54	2134.53	2681.22	2041.73
Concrete Data	11243.49	6806.26	7145.05	7185.30	7439.98	7096.97
Bike Sharing	607563.72	416858.08	429205.43	411248.85	436794.38	408658.06

Diversity and QBC algorithms are statistically significantly superior to the random baseline at $p < 0.05$.

Figure 2 shows the total time spent waiting for each selection strategy to issue label queries. Increased batch sizes are shown to significantly reduce the time taken for the execution time for each selection strategy. We include these as an indicator of the time *cost* of employing each selection strategy. Generally speaking, the slower the execution time, the more time an oracle spends waiting for queries, and the greater the actual or opportunity cost of employing active learning [24]. Although QBC and Diversity have broadly similar execution times, it has been noted that as the Diversity strategy in no way depends on the actual labels provided by an oracle, it is possible to pre-compute the label requests prior to beginning the active learning loop [7]. Implementing the Diversity selection strategy to take advantage of this

Table 3 Ranked performance for all algorithms, with the lowest rank for each dataset shown in bold

	Density	Diversity	EGAL	EGL	Mean Random	QBC	Total
Auto MPG	6	1	5	3	4	2	21
Wine Quality Red	6	1	4	3	5	2	21
Wine Quality White	6	4	3	1	5	2	21
Treasury	6	1	4	3	5	2	21
House Prices	6	3	4	2	5	1	21
Concrete Data	6	1	3	4	5	2	21
Bike Sharing	6	3	4	2	5	1	21
Total	42	14	27	18	34	12	147

Table 4 Pairwise comparisons using Nemenyi multiple comparison test with q approximation for unreplicated blocked data

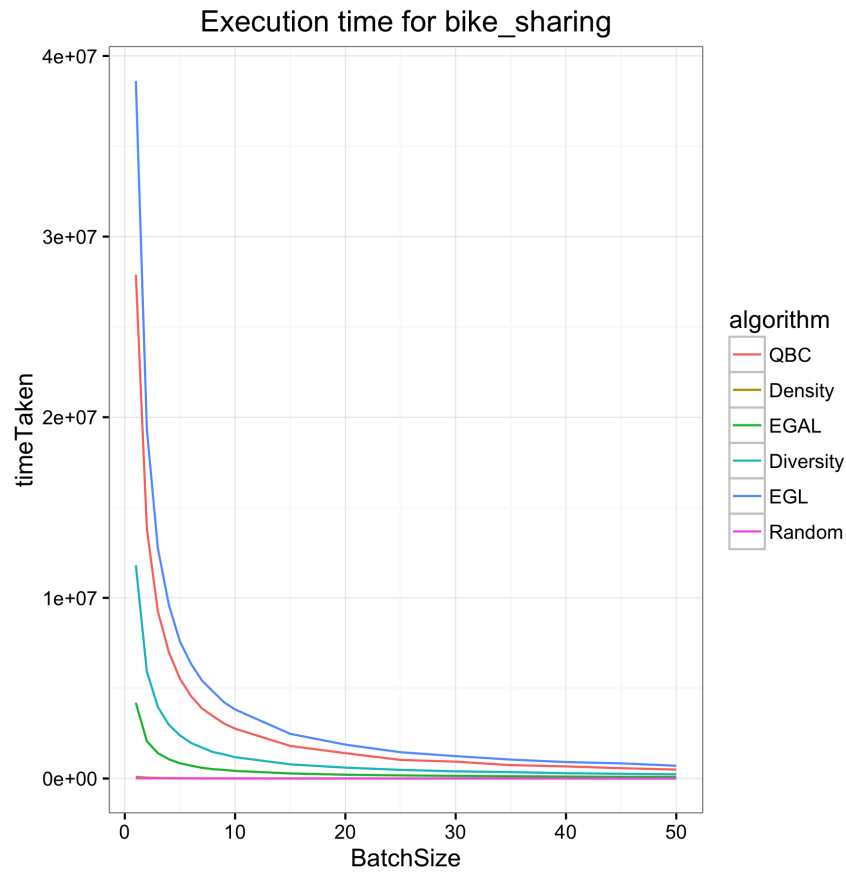
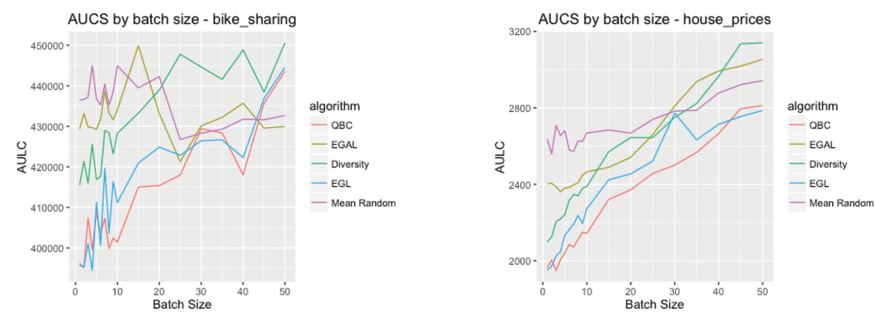
	Density	Diversity	EGAL	EGL	Random
Diversity	0.00090				
EGAL	0.26521	0.42892			
EGL	0.00798	0.99287	0.79304		
Random	0.86345	0.04896	0.91810	0.19978	
QBC	0.00026	0.99975	0.26521	0.95648	0.02076

would eliminate the waiting period between batches, making it a more attractive option in real-world scenarios.

Although increased batch sizes have a positive effect on the time cost of each selection strategy, this is balanced by a negative impact on the model’s overall performance. The Diversity and QBC algorithms outperformed the Random baseline at batch sizes up to and including 10. The relative performance of the selection strategies decreased dramatically as batch size increased. Figure 3 demonstrates the impact of increasing batch size on each selection strategies’ AULC. Real world application making use of active learning will therefore have to strike a balance between speed of execution and accuracy of predicted labels.

5 Conclusion

In this paper we evaluated a number of active learning approaches for regression, including the existing state-of-the-art for classification problems. The QBC selection strategy, which has been shown to perform well in classification problems, maintained a consistent significant improvement over a random baseline in regression problems. Interestingly, a purely diversity-based approach, which has received lit-

Fig. 2 Execution time on the bike sharing dataset by batch size**Fig. 3** Selection Strategy performance on selected datasets by batch size

the attention in classification problems, was shown to significantly outperform the baseline.

The choice of batch size was found to be an important factor in determining both the overall accuracy and execution time of both model-based and model-free selections strategies. A higher batch size reduces execution time but also negatively impacts selection strategy performance. In the case of model-free selection strategies, the impact of increased execution time can be reduced by generating batch queries before the active learning loop begins.

The strong performance of the Diversity algorithm suggests that further research into active learning strategies based solely on integral properties of the dataset shows promise in developing algorithms which are not only more computationally efficient, but also more accurate than the more popular ensemble-based alternatives. In particular, the EGAL selection strategy may be improved by adjusting the balancing parameter to put more emphasis on diversity measures when used in the context of a regression problem.

While Diversity was shown to perform well in a data labelling task, it remains to be seen whether models induced using these selection strategies can generalize well to predicting unseen data. A natural extension of the current study would be to examine the effectiveness of these approaches within this common scenario.

References

1. B. Settles, "Active learning literature survey," *University of Wisconsin, Madison*, vol. 52, no. 55-66, p. 11, 2010.
2. B. Settles, M. Craven, and L. Friedland, "Active learning with real annotation costs," in *Proceedings of the NIPS workshop on cost-sensitive learning*, pp. 1–10, 2008.
3. D. D. Margineantu, "Active cost-sensitive learning," in *IJCAI*, vol. 5, pp. 1622–1623, 2005.
4. R. Hu, S. J. Delany, and B. Mac Namee, "Egal: Exploration guided active learning for tcbr," in *Case-Based Reasoning. Research and Development*, pp. 156–170, Springer, 2010.
5. W. Cai, Y. Zhang, and J. Zhou, "Maximizing expected model change for active learning in regression," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pp. 51–60, IEEE, 2013.
6. R. M. Castro, R. Willett, and R. Nowak, "Faster rates in regression via active learning," in *Advances in Neural Information Processing Systems*, pp. 179–186, 2005.
7. K. K. Sung and P. Niyogi, "Active learning for function approximation," *Advances in neural information processing systems*, pp. 593–600, 1995.
8. K. Yu, J. Bi, and V. Tresp, "Active learning via transductive experimental design," in *Proceedings of the 23rd international conference on Machine learning*, pp. 1081–1088, ACM, 2006.
9. H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the fifth annual workshop on Computational learning theory*, pp. 287–294, ACM, 1992.
10. R. Burbidge, J. J. Rowland, and R. D. King, *Active learning for regression based on query by committee*, pp. 209–218. Springer, 2007. state-of-the-art.
11. B. Settles, M. Craven, and S. Ray, "Multiple-instance active learning," in *Advances in neural information processing systems*, pp. 1289–1296, 2008.
12. Y. Baram, R. El-Yaniv, and K. Luz, "Online choice of active learning algorithms," *The Journal of Machine Learning Research*, vol. 5, pp. 255–291, 2004.
13. T. Osugi, D. Kim, and S. Scott, "Balancing exploration and exploitation: A new algorithm for active machine learning," in *Data Mining, Fifth IEEE International Conference on*, pp. 8–pp, IEEE, 2005.
14. E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526–558, 2009.
15. B. Mac Namee, R. Hu, and S. Delany, "Inside the selection box: Visualising active learning selection strategies," in *The Challenges of Data Visualization Neural Information Processing Systems, (NIPS) 2010 Workshop*, 2010.
16. R. Hu, *Active Learning For Text Classification*. Thesis, Dublin Institute of Technology, 2011.
17. B. Settles, "From theories to queries: Active learning in practice," *Active Learning and Experimental Design W*, pp. 1–18, 2011.
18. Z. Xu, R. Akella, and Y. Zhang, *Incorporating diversity and density in active learning for relevance feedback*. Springer, 2007.
19. S. C. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proceedings of the 23rd international conference on Machine learning*, pp. 417–424, ACM, 2006.
20. Y. Guo and D. Schuurmans, "Discriminative batch mode active learning," in *Advances in neural information processing systems*, pp. 593–600, 2008.
21. R. Hu, B. Mac Namee, and S. J. Delany, "Sweetening the dataset: Using active learning to label unlabelled datasets," in *Proceedings of the 19th. Irish Conference on Artificial Intelligence and Cognitive Science (AICS '08)*, 2008.
22. T. RayChaudhuri and L. G. Hamey, "Minimisation of data collection by active learning," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, vol. 3, pp. 1338–1341, IEEE, 1995.
23. H. W. Borchers, *pracma: Practical Numerical Math Functions*, 2015. R package version 1.8.6.

24. R. Haertel, P. Felt, E. Ringger, and K. Seppi, "Parallel active learning: eliminating wait time with minimal staleness," in *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pp. 33–41, Association for Computational Linguistics, 2010.